DOCUMENT RESUME

ED 220 508 TM 820 523

| | |
|---|---|
| AUTHOR | Arter, Judith A.; And Others |
| TITLE | State-of-the-Art in Evaluation Reports and in Selecting Exemplary Reports: AERA Competition for Best Report of an Evaluation. |
| PUB DATE | Mar 82 |
| NOTE | 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982). |
| EDRS PRICE | MF01/PC01 Plus Postage. |
| DESCRIPTORS | Competitive Selection; Decision Making; *Evaluation Criteria; Evaluation Methods; Guidelines; Rating Scales; *Research Reports; *Standards; *State of the Art Reviews; *Writing Evaluation |
| IDENTIFIERS | American Educational Research Association; *Evaluation Reports; Northwest Regional Educational Laboratory |

ABSTRACT

The Northwest Regional Educational Laboratory was involved in selecting the "Best Report of an Evaluation" for the 1981 awards competition of Division H of the American Educational Research Association (AERA). During this process two things became apparent--there was no procedure or set of standards by which to select a best report, and the quality of reports submitted was highly variable. A procedure and a set of criteria were devised for rating the reports. These should prove useful for providing local report writers a simple set of guidelines for preparing reports that have a greater chance of aiding decision making, and for standardizing the AERA procedure for rating reports. This paper describes both the rating scale and the characteristics of the reports examined. In general it was concluded that evaluation reports need to be clearer on the purpose and audience for the evaluation, on specific evaluation questions, and on sound and useful suggestions for program improvement. Also, evaluation reports need to contain a short, but useful, summary which covers the questions of major concern to the audience of the evaluation. (Author/PN)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. A. Arter
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED220508

# State-of-the-Art in Evaluation Reports and in Selecting Exemplary Reports:

## AERA Competition for Best Report of an Evaluation

Judith A. Arter, Randy Nelson, Joe B. Hansen, Warren Evans
Northwest Regional Educational Laboratory

## Background

In January of 1980 NWREL staff were asked to assist in selecting the "Best Report of an Evaluation" for the 1981 annual awards competition. There are two purposes for this paper.

First, we will report on our development of a rating scale and the iterative procedure used to select the best report as a contribution to standardizing the awards competition. AERA initiated its awards competition in recognition of the growing emphasis on evaluation for decision making. Even though the competition has been going on for some years, there is no set of procedures or standards by which to judge entries. Even though the expertise of the judges probably does result in choosing the best report, it would be useful to establish a set of standards which would provide guidance for submitters and provide consistency in judging across years. These standards could be nothing more than a short list of criteria by which reports are judged.

The second purpose of the paper is to report on the state-of-the-art in evaluation report writing. There have been numerous attempts to provide guidelines for writing evaluation reports (see Appendix A). Even so, many evaluation reports still lack much of the content and structure that would make them easily accessible for decision making. This paper discusses the characteristics of almost 50 evaluation reports submitted for the AERA awards competition in light of their usefulness for decision making.

TM 820 523

Paper presented at the annual meeting of the American Educational Research Association, March 19-23, 1982, New York, New York.

## Procedures

A committee of five persons was convened to judge the evaluation reports submitted for the 1981 competition. Using existing discussions of the qualities of an evaluation report that makes it useful for decision making (see Appendix A), a draft of a rating checklist was devised and reviewed by the committee. Based on a field test, the checklist was revised for clarity and comprehensiveness.

The final rating form consisted of 20 items ranging from a report format that allowed for easy access to information, to the writing style of the report, to the technical adequacy of the report. Each report was rated on a scale of one to five for each item. Total scores could thus range from 20 to 100. Definitions were provided for each item to provide raters with a common understanding of how to rate reports. (The rating form and definitions is included as Appendix B.)

The process for choosing the best report consisted of three stages. First, the evaluation reports were randomly assigned to raters. The raters judged each report and nominated the best one or two for further consideration. The reports that survived the first cut were randomly assigned to a second rater. The totals from the two ratings were calculated and the top three reports were rated by a single third person. The winner was chosen from the three finalists.

This process had several advantages--interrater reliability problems were minimized by a series of iterations, and more than one rating was involved in choosing the winner.

## Properties of the Rating Form

Several analyses were run in order to get a perspective on whether the rating form was adequate for the purpose for which it was used--a sequential series of selecting the top evaluation reports in order to arrive at the single "best" report. We did not expect the instrument to have text-book measurement properties since it did not receive a rigorous development and review process. Therefore, it probably would not be useful for research situations in which one was interested in actual comparisons between groups of ratings. However, the instrument might still be useful for the selection process described in this paper.

### Approximation to Interrater Reliabilities

Only nine of the original 46 evaluation reports received ratings by more than one individual--four received two ratings and five received three ratings. This was because the selection process used to identify the best evaluation report called for selecting the best reports at each step. These best reports then went on to the next round of selection. Because of this process, true interrater reliabilities were not possible. Our approximation to this was to calculate each rater's average rating on each item and to see how close these average rankings were. If the average ratings of all five raters did not differ by more than one point, the consistency of ratings was called "good". A difference of 1.0 to 1.5 was called "fair", and a difference of over 1.5 was called "poor". This process was considered justifiable because evaluation reports were randomly distributed to raters, and each rater did at least ten reports.

Table 1 shows the average rating on each item by each rater. Using the criteria described above, the ratings on seven items are "good", the ratings on eleven items were "fair", and the ratings on two items were "poor." Items 18 and 19 were of the most concern. These items dealt with whether recommendations in the report were justifiable (based solidly on the data presented in the report) and useful to the audience. The raters differed a great deal in their ratings on these items--perhaps more than can be explained by random differences in the quality of the reports assigned to each rater.

The raters may have used different criteria on which to judge these items. For example, a judgement on the usefulness of the recommendations depends on whether an audience for the evaluation is specified. Some raters may have given a low rating to item 19 solely because the audience was not specified, and therefore one could not tell whether the recommendations were useful. Others, in this situation, may have used a "default" audience against which to judge usefulness. Perhaps the criteria for rating these items needs to be more clearly specified.

The items most consistently rated were items two and ten--ease of finding information and the adequacy of the evaluation design. The latter is perhaps not surprising given that technical persons were rating the reports.

Item-Total Correlations

To get an idea of how the items were contributing to the total score we calculated item-total score correlations. These are presented in Table 2. Three correlations were below .26, seven were between .27 and .49, five were between .50 and .59 and five were .60 or greater.

Arter, J.A., Nelson, R., Hansen, J., & Evans, W. State-of-the-Art in Evaluation Report Writing. AERA, 1982

Table 1

Average Ratings for Each Rater

| Item | Joe (N=11) | Judy (N=16) | Randy (N=12) | Barry (N=10) | Warren (N=11) | | Total (N=60) |
|---|---|---|---|---|---|---|---|
| 1 | 4.8 (.60)* | 3.8 (1.17) | 3.9 (1.0) | 4.7 (.5) | 4.4 (.8) | F | 4.3 (.95) |
| 2 | 4.0 (1.41) | 4.1 (1.20) | 4.2 (1.2) | 3.9 (1.0) | 3.9 (1.4) | G | 4.0 (1.22) |
| 3 | 3.6 (1.50 ) | 2.6 (1.26) | 3.1 (1.5) | 3.1 (1.9) | 3.3 (1.3) | F | 3.1 (1.47) |
| 4 | 2.4 (1.29) | 2.9 (1.44) | 3.3 (1.3) | 3.7 (1.2) | 3.6 (1.2) | F | 3.2 (1.34) |
| 5 | 3.6 (1.50) | 3.9 (1.57) | 3.3 (1.6) | 4.3 (1.3) | 3.8 (1.2) | G | 3.8 (1.45) |
| 6 | 3.5 (1.57) | 3.7 (1.49) | 3.4 (1.6) | 4.1 (1.3) | 4.0 (1.2) | G | 3.7 (1.42) |
| 7 | 2.7 (1.62) | 3.1 (1.61) | 2.8 (1.0) | 3.9 (1.0) | 3.9 (.9) | F | 3.2 (1.36) |
| 8 | 3.1 (1.38) | 4.1 (1.1) | 4.1 (.8) | 3.6 (1.1) | 3.6 (.9) | F | 3.7 (1.09) |
| 9 | 3.5 (.93) | 3.9 (1.4) | 4.1 (.3) | 3.5 (1.1) | 4.1 (.8) | G | 3.8 (.99) |
| 10 | 3.7 (.65) | 3.3 (1.5) | 3.6 (.7) | 3.7 (1.2) | 3.7 (1.0) | G | 3.6 (1.06) |
| 11 | 4.3 (.90) | 3.6 (1.4) | 3.7 (.9) | 3.3 (1.2) | 4.2 (.9) | G | 3.8 (1.14) |
| 12 | 3.7 (.65) | 3.1 (1.5) | 3.9 (.9) | 3.8 (.9) | 4.3 (.6) | F | 3.7 (1.08) |
| 13 | 3.5 (.93) | 3.8 (1.2) | 4.0 (1.0) | 4.2 (.8) | 4.3 (.9) | G | 3.9 (.99) |
| 14 | 2.9 (1.58) | 3.5 (1.6) | 4.2 (.9) | 4.2 (1.0) | 4.2 (1.2) | F | 3.8 (1.38) |
| 15 | 4.7 (.47) | 3.7 (1.1) | 3.7 (1.4) | 3.5 (.5) | 4.5 (.7) | F | 4.0 (1.03) |
| 16 | 3.4 (1.21) | 3.7 (1.5) | 4.0 (.7) | 4.3 (.7) | 4.0 (1.3) | G | 3.8 (1.18) |
| 17 | 3.4 (1.81) | 4.1 (1.2) | 4.2 (1.1) | 4.5 (.5) | 3.6 (1.3) | F | 4.0 (1.27) |
| 18 | 1.9 (1.58) | 2.8 (1.8) | 3.4 (1.2) | 4.1 (1.1) | 3.3 (1.6) | F | 3.0 (1.62) |
| 19 | 1.8 (1.40) | 2.6 (1.2) | 3.6 (1.8) | 4.1 (1.1) | 3.1 (1.4) | F | 3.0 (1.55) |
| 20 | 3.3 (.65) | 3.4 (1.5) | 4.1 (1.0) | 3.6 (.5) | 4.3 (.6) | F | 3.7 (1.03) |
| Total | 68.0 (9.97) | 69.6 (14.2) | 74.4 (10.1) | 78.1 (7.99) | 78.1 (11.22) | | 73.25 (11.64) |

* The numbers in parentheses are standard deviations.

Table 2

Correlations Between Item Ratings and Total Scores

| Item | Correlation (N=60) |
|---|---|
| 1. Cover page | .25 |
| 2. Ease of finding information | .40 |
| 3. Adequate summary | .34 |
| 4. Purpose and intended audience | .61 |
| 5. Clearly stated program goals | .33 |
| 6. Adequate project description | .42 |
| 7. Clear, measurable evaluation questions | .52 |
| 8. Appropriate description of methodology | .45 |
| 9. Adequate procedure | .55 |
| 10. Adequate evaluation design | .60 |
| 11. Adequate instruments | .54 |
| 12. Necessary and sufficient analyses | .53 |
| 13. Appropriate description of results | .33 |
| 14. Results clearly related to questions | .71 |
| 15. Tables, charts and graphs | .14 |
| 16. Appropriate interpretation of findings | .62 |
| 17. Clear statement of conclusions | .54 |
| 18. Sound suggestions for program improvement | .70 |
| 19. Useful suggestions for program improvement | .45 |
| 20. Appropriate writing style | .26 |

The items correlating lowest with the total scores were item 1 (cover page), item 15 (clarity of tables and graphs), and item 20 (appropraite writing style). The items correlating the highest with the total score were item 4 (purpose and itended audience specified), item 10 (adequacy of the evalution design), item 14 (results related to questions in a clear fashion), item 16 (appropriate interpretations of results) and item 18 (recommendations based solidly on the data). In the usage of the instrument it appears that technical concerns contributed more to positive judgements about the quality of an evaluation report than did concerns about communication.

This may be because there was a restriction in range on style in the reports submitted, it might reflect the bias of the judges, or it might be that communication style does not differentiate good and bad reports. In any case, we need to seriously consider two issues. First, which area should be considered most important to the "best report of an evaluation"--technical soundness of the evaluation or the manner in which it communicates to audiences? Secondly, is it possible to separate the two? That is, is it possible to have an evaluation report that communicates effectly if the information on which the report is based is not of good quality? It was the opinion of the committee that communication items should not be deleted from the checklist.

Factor Analysis

Since the rating form was developed to reflect at least two independent dimensions (ability to communicate and technical soundness) of evaluation reports, we decided to factor analyze the results to see whether these two factors emerge. For purposes of the factor analysis we used only the first round of ratings. It was thought that adding second

and third ratings for the same reports might weight the correlations for the better reports. The technique used was specified in SPSS as principal factoring with iteration. We generated both orthogonal and oblique factor loadings.

The results showed six factors accounting for 71% of the variance. Table 3 shows the communalities and factor matrix of the orthogonal factor analysis. Table 4 describes the six factors. The factor accounting for the most variance (25.6%) in the ratings is a technical factor. This loads on item 7 (clear evaluation questions), item 10 (adequate design), item 12 (adequate analyses), item 11 (adequate instruments), item 9 (adequate procedure), and item 14 (results related to questions).

The next four factors contain factors that, for the most part, fit logically together. It is reasonable that these factors would be orthogonal--since it is possible to see how ratings on one set of items would be independent of ratings on another set of items. These four items are the adequacy of the recommendations, the presence of a program description, ease of finding information in the report, and a appropriate writing style. The sixth factor is difficult to interpret.

We do not wish to dwell on the factor analysis, since we feel that the ratings may not warrent such detailed examination. However, we feel reasonably assured that the structure and functioning of the instrument is sufficient for the purpose intended. An issue that should be addressed relates again to the fact that the technical adequacy of the evaluation reports weighs very heavily on decisions as to the quality of the report. Again, one needs to decide whether this should be the case.

Table 3

Factor Analysis

| Item | Factor Matrix | | | | | | Communalities |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | |
| 1. Cover page | .02 | -.11 | .24 | .09 | -.14 | .57 | .42 |
| 2. Ease of finding information | .21 | .08 | .06 | .68 | -.04 | -.10 | .53 |
| 3. Adequate summary | .02 | .06 | -.14 | .54 | .06 | .30 | .41 |
| 4. Purpose and intended audience | .32 | .43 | .27 | -.08 | .12 | .44 | .57 |
| 5. Clearly stated program goals | -.06 | .20 | .82 | -.22 | .02 | .21 | .81 |
| 6. Adequate project description | .11 | .09 | .69 | .15 | .08 | .05 | .53 |
| 7. Clear, measurable evaluation questions | .63 | .09 | -.34 | -.02 | -.01 | .42 | .40 |
| 8. Appropriate description of methodology | .42 | .35 | .14 | .18 | -.07 | -.21 | .40 |
| 9. Adequate procedure | .80 | .06 | .07 | .13 | -.12 | -.18 | .71 |
| 10. Adequate evaluation design | .86 | -.10 | .04 | .09 | .04 | .10 | .77 |
| 11. Adequate instruments | .69 | -.09 | .13 | .15 | .02 | .04 | .53 |
| 12. Necessary and sufficient analyses | .76 | -.04 | .06 | .05 | .35 | .15 | .73 |
| 13. Appropriate description of results | .08 | .03 | -.03 | -.04 | .71 | -.01 | .52 |
| 14. Results clearly related to questions | .74 | .40 | -.14 | -.04 | .12 | .23 | .79 |
| 15. Tables, charts and graphs | .27 | -.52 | -.06 | -.01 | .24 | .32 | .51 |
| 16. Appropriate interpretation of findings | .43 | .27 | -.02 | .21 | .23 | -.07 | .36 |
| 17. Clear statement of conclusions | .13 | .44 | .07 | .56 | -.18 | -.03 | .57 |
| 18. Sound suggestions for program improvement | .19 | .75 | .24 | .26 | .12 | .08 | .75 |
| 19. Useful suggestions for program improvement | -.04 | .85 | .06 | .10 | .23 | -.01 | .80 |
| 20. Appropriate writing style | .04 | .14 | .31 | -.04 | .46 | -.08 | .34 |
| Eigenvalues | 5.1 / 4.8 | 2.9 / 2.5 | 2.0 / 1.6 | 1.6 / 1.1 | 1.4 / 0.9 | 1.3 / 0.8 | |
| Percent of Variance | 26 / 41 | 14 / 22 | 10 / 14 | 8 / 10 | 7 / 8 | 6 / 7 | |

Table 4

Orthogonal Factors

| Factor | Interpretation |
|---|---|
| 1 — Technical | Clear evaluation questions, adequate evaluation design, adequate analyses, adequate instruments, adequate procedure, results related to questions. |
| 2 — Recommendations | Recommendations based soundly on the evaluation results, recommendations useful for the audience of the evaluation. Some loading on purpose and audience of evaluation, results related to questions, appropriate tables and clear conclusions. |
| 3 — Program Description | Adequate description of the program, adequate specification of program goals. |
| 4 — Ease of Finding Information | Ease of finding information, presence of an adequate summary, clear conclusions. |
| 5 — Appropriate Writing Style | Appropriate writing level in the description section. Some loading on overall writing style. |
| 6 — Uninterpretable | |

Table 1 can also be used to examine the overall strengths and weaknesses of the reports submitted to the AERA competition. First, however, it is important to note that the sample of reports submitted to the competition is probably not representative of the universe of reports. Whether they are better or worse than most is not clear, but the reports submitted revealed some severe weaknesses that limit their potential usefulness.

Focusing on the average ratings for each item across all raters, the weakest areas of the reports in general were on item 3 (adequate summary), item 4 (statement of the purpose of the evaluation and the intended audience), item 7 (presence of clear, measurable evaluation questions), item 18 (sound suggestions for program improvement, and item 19 (useful suggestions for program improvement). The average rating for these items all ranged from 3.0 to 3.2 with standard deviations from 1.34 to 1.62. Subjective reactions of the raters to the evaluation reports were that the quality was generally poor with major problems being triviality, poor design and exclusion of important information.

The strongest areas of the reports, in general, were item 1 (adequacy of the title page), item 2 (ease of finding information), item 15 (adequacy of tables, charts and graphs), and item 17 (clear statement of conclusions). These items all had average ratings from 4.0 to 4.3 with standard diviations from .95 to 1.27. These judgements were generally based on the fact that tables of contents were detailed, that tables and charts were clear and were discussed in the text, and that conclusions were present.

It might be helpful in improving evaluation reports to generate and disseminate standards by which reports are judged in the competition. This might provide an impetus to report writers to make their reports more useful.

## Conclusions

There is a wide variety of opinion as to what should be included in an evaluation report. Even the raters tended to look for and weigh different things based on background and interest. We feel, however, that the rating form and procedure used in this study is adequate for use as a general and quick guide to the important content of an evaluation report, and is a useful way to judge the quality of evaluation reports. We feel that this or a similar set of standards should be adopted by AERA in order to inform entrants of the criteria used to judge reports, in order to provide more consistency in the process across years, and to provide guidance to report writers as to what things should be included on an evaluation report.

From the study of reports, it appears that many current evaluation reports have serious deficits which detract from their usefulness for making decisions. In general, evaluation reports need to be clearer on the purpose and audience for the evaluation, on specific evaluation questions, and on sound and useful suggestions for program improvement. Also, evaluation reports need to contain a short, but useful summary which covers the questions of major concern to the audience of the evaluation.

## APPENDIX A

### LITERATURE REVIEW ON CHARACTERISTICS OF DESIRABLE EVALUATION REPORTS

In preparing the rating form, the committee reviewed several books and papers which discussed the desirable content and style of an evaluation report. The four sources represent a sampling of the many references available, and those cited in this paper were chosen because of their completeness and familiarity to evaluators. The order in which they will be reviewed is chronological, and does not reflect differences in content or import. In fact, all four sources were largely in agreement on the characteristics of a good evaluation report. They differed mainly in the emphasis each author placed on the major facets of the report.

The U.S. Office of Education's Preparing Evaluation Reports: A Guide for Authors (1970) identifies five major parts of an evaluation paper. They are (1) the summary, (2) a description of the context, (3) an explanation of the program, (4) the reporting of the evaluation and (5) the recommendations. For each part of the report, the guide lists and discusses questions which may be answered in an evaluation report.

After a cursory description of the summary, a complete and detailed discussion of program context is given. Included in the discussion are population and economic patterns of the locale, the organization and financial status of the school system and special factors such as needs assessment findings and other historical background. The secion on program description was the most extensive of any of the references reviewed. Program objectives and participants, personnel, budget and implementation procedures were allotted 20 pages of the guide. Most

detailed was the discussion of project implementation which included an examination of organizational details, project activities and services, instructional equipment and materials and parent-community involvement.

The segment on evaluation procedures and results was brief and condensed. Topics covered included program objectives, participants, measuring change, presenting and analyzing data and report findings. The guide concludes with a brief discussion on preparing recommendations. Coverage of the more technical topics associated with evaluation is not attempted, but additional references in research design, measurement and statistics are suggested.

In Worthen and Sanders' Educational Evaluation: Theory and Practice (1973), the insight and experience of numerous authors are called upon when discussing evaluation reporting. Three distinct sets of helpful information are offered in the chapter entitled "Considerations in Planning Evaluation Studies." The first is a listing of criteria for judging evaluation studies borrowed from Stufflebeam, et al. (1971). Three groups of criteria, labeled scientific, practical and prudential are briefly outlined. Scientific criteria refers to the internal and external validity of an evaluation as well as the reliability and objectivity of the information collected. The relevance, importance, scope, credibility, timeliness and persuasiveness of an evaluation constituted the practical criteria. The efficiency of an evaluation was referred to as the prudential criterion. Although these criteria were not mentioned specifically within the context of evaluation reporting, they are standards which should be reflected in a complete evaluation report.

Worthen and Sanders later present their own suggested format for an evaluation proposal. Although it is not an outline for a final report, all the elements listed in the proposal are relevant and are worthy of inclusion in the final report.

The five major sections of their suggested proposal are (1) the rationale, (2) the objectives of the evaluation study, (3) the description of the program, (4) the evaluation design and (5) a description of the final report. The points to be covered under each of these subheadings are sound and fairly standard. The outline does differ from most, however, in its emphasis on evaluator bias. In the proposal, the evaluators' philosophy about the evaluation is to be described which includes conscious biases which may affect the reported outcomes.

The third section of interest in the Worthen and Sanders book is a paper by Stake entitled "Evaluation Design, Instrumentation, Data Collection and Analysis of Data" (1969). Stake offers a general table of contents for a final evaluation report. The five sections in the report are (1) objectives of the evaluation, (2) specification of the program, (3) program outcomes, (4) relationships and indicators and (5) judgments of worth. When discussing the objectives of an evaluation, the first consideration is the audience served by the evaluation. Different audiences will require different information. He then suggests that the evaluator discuss the possible decisions that can be made about the program as a result of the evaluation. The second section describes the implementation and processes of the program. A relatively high emphasis is placed on the discussion and description of the educational philosophy of the program, subject matter, learning objectives and procedures, students and instructional context. The section on program outcomes

includes coverage of the opportunities and experiences provided, student gains, side effects and all related costs. The next section examines the outcomes of the evaluation in terms of what was expected and what caused the results to be what they were. The report would conclude with an examination of the worth or value of the program, and of the evaluation. This evaluation report outline was far different in its structure and emphasis than most of the others viewed. It placed far less importance on the technical and methodological aspects of the evaluation and relied far more on the evaluator's ability to describe and document the personal and social effects of a program.

Morris and Fitz-Gibbon devote an entire chapter to "An Evaluation Report Outline" in their book How to Present an Evaluation Report (1978). The outline contains seven sections plus specifications for the information to be found on the front cover of the report. The sections recommended for the report are (1) the summary, (2) background information, (3) description of the evaluation, (4) results, (5) discussion of the results, (6) costs and benefits and (7) conclusions and recommendations. Under each section heading is a detailed list of the issues to be addressed. The treatment is complete, but fairly standard.

Of greater interest in the discussion of communication of evaluation esults. After mentioning several points having to do with audience identification, the authors suggest that to be effective, a report should not give an audience more information than it needs. The chapter goes on to discuss how to highlight important points and how to write in a readable style. There is also a lengthy chapter on how to present graphs and tables. Thus, this book differs from the others previously mentioned

in its greater emphasis on style. The clarity and attractiveness of a report, while less important that its content, certainly is a major contributor to the effectiveness of an evaluation report.

The last document to be mentioned in this brief review does not deal directly with the contents of an evaluation report, but rather it outlines standards which should be met in all educational evaluations. The Standards for Evaluations of Educational Programs, Projects and Materials (1981) were developed by a multitude of persons referred to as the Joint Committee on Standards for Educational Evaluation. Thirty standards are identified which are presented in four categories. Only selected standards will be mentioned as some have been already addressed by other sources and others are not particularly relevant to evaluation reporting.

The standards relating to the utility of an evaluation include audience identification, valuation interpretation, report clarity and other issues already discussed. They also include, however, (1) report timeliness and (2) evaluation impact. Timeliness means that a report should be presented so that the information may be best used. Impact refers to an evaluation which encourages follow-through by members of the audience. These are both often difficult to assess when examining a report, but are nonetheless quite important.

The feasibility standards are intended to ensure that evaluations are prudent and reasonable. Although the standards do not refer directly to the evaluation report, one implication might be that reports should not be overly extravagant in their production.

The propriety standards are designed to ensure that proper legal and ethical conduct is maintained. Two standards directly relate to

evaluation reports. The first, full and frank disclosure, requires that reports be direct and honest in their presentation of results. The second, balanced reporting, requires that a complete and fair presentation of program strengths and weaknesses be given. As is the case with the utility standards, whether these standards have been met is often difficult to assess with no other information available other than the completed final report.

The accuracy standards were designed to ensure the technical quality of an evaluation. Many of the standards have been mentioned by previously cited sources. One which deserves special mention is the standard for justified conclusions. This implies that all conclusions cited in an evaluation report should be derived directly from the results of the evaluation.

Reviewing these materials made several points clear. The suggested organization and content of an evaluation report by various authors, although very similar in many respects, reflects the evaluator's values and biases about the nature of the evaluation. Secondly, it seems unlikely that any report could satisfy all criteria of all the suggested outlines. Thus, the committee attempted to take a balanced approach which sampled concerns from all the approaches.

Report No. _____
Total Points _____
Rater _____

Appendix B

## EVALUATION REPORT RATING FORM

Please rate each evaluation report on each of the following 20 criteria, using a five point system indicating the degree to which the criterion of interest has been satisfied.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| Highly Unsatisfactory | | | | Highly Satisfactory |

1. Cover page _____
2. Ease of finding information _____
3. Adequate summary _____
4. Purpose and intended audience _____
5. Clearly stated program goals _____
6. Adequate project description _____
7. Clear, measurable evaluation questions _____
8. Appropriate description of methodology _____
9. Adequate procedure _____
10. Adequate evaluation design _____
11. Adequate instruments _____
12. Necessary and sufficient analyses _____
13. Appropriate description of results _____
14. Results clearly related to questions _____
15. Tables, charts and graphs _____
16. Appropriate interpretation of findings _____
17. Clear statement of conclusions _____
18. Sound suggestions for program improvement _____
19. Useful suggestions for program improvement _____
20. Appropriate writing style _____

## EVALUATION REPORT RATING FORM
### Description of Items

1. Does the report cover page bear all the information needed for a reader to obtain further information such as: title, author(s), agency or contractor, date, etc. ?

2. Can information be easily found in the report--logical chapters and subheadings, adequate table of contents and list of figures?

3. Is there an adequate summary--purpose, questions, major findings, recommendations?

4. Is there a statement which adequately describes the purpose of the evaluation and the intended audience?

5. Are the objectives or goals of the program, product or process clearly stated? If not, is there an adequate explanation for the deficiency?

6. Does the report include sufficient description of project context, background and process?

7. Are the evaluation questions clearly specified in measurable terms?

8. Is the description of evaluation methodology (procedure, design, instruments and analyses) appropriate for the audience?

9. Is the procedure adequate (for example, quality control in testing and handling data, timelines, ethics, etc.)?

10. Does the evaluation design include adequate controls for threats to internal validity (maturation, history, selection, testing, instrumentation, regression, mortality, diffusion of the treatment, etc.)?

11. Are the instruments adequate to answer the evaluation questions--are they reliable and valid for the purposes intended?

12. Are the analyses necessary and sufficient to answer the evaluation questions?

13. Is the description of results appropriate for the audience?

14. Are the results related to the questions of interest in a clear fashion?

15. Are tables, charts and graphs used appropriately to illustrate the results--clearly labelled, appropriate for the audience, containing no information not discussed in the text?

16. Has caution been evidenced in interpreting the findings so as to not draw inferences beyond the range of the data and procedures used?

17. Are conclusions stated clearly and unequivocally?

18. Are suggestions offered for program improvement based soundly on the results?

19. Are suggestions offered for program improvement useful--meaningful, practical, timely, etc.?

20. Is the writing style of the report appropriate for the intended audience?